

блоку P06 через порт COM3 по интерфейсу RS–485 подключены 176 интеллектуальных датчиков температуры ТСТ11 для контроля контуров охлаждения. Существующие датчики с аналоговым и дискретным выходным сигналом подключаются к модулям ввода-вывода ТЕКОНИК® (до 10 типов модулей для различных типов датчиков и исполнительных механизмов), а сами модули формируют цифровую сеть через порт COM2 с интерфейсом RS – 485. Далее технологические данные поступают в существующую АСУ ТП.

Использование ТСТ11 позволяет повысить точность измерений и получить существенную экономию на монтажных работах. Значительно облегчается создание температурного контроля в случаях, когда затруднена прокладка кабельных трасс для традиционных датчиков, что особенно характерно для рассматриваемой системы.

Список использованных источников

1. Ванюков А.В. Плавка в жидкой ванне. – М.: Металлургия, 1988. – 207 с.
2. Кривандин В.А. Теплотехника металлургического производства. Т. 2. Конструкция и работа печей: Учебное пособие для вузов / Под ред. В.А. Кривандина]. – М.: МИСИС, 2001. – 736 с.
3. Филимонов Д.А. Система интеллектуальных модулей ТЕКОНИК® [Электронный ресурс] // Отраслевой научно-технический журнал «ИСУП». – 2007. – Режим доступа: <http://isup.ru/articles/4/264>.
4. Шишов О.В. Интеллектуальные датчики в системах промышленной автоматизации [Электронный ресурс] // Электронное научное издание "Электроника и информационные технологии". – 2011, выпуск 2 (11). – Режим доступа: <http://fetmag.mrsu.ru>.

УДК 004-04

В. И. Переплетчиков, В. Ю. Носков

ФГАОУ ВО «Уральский федеральный университет

имени первого Президента России Б.Н. Ельцина», г. Екатеринбург, Россия

РАЗРАБОТКА ИНФОРМАЦИОННОЙ СИСТЕМЫ СБОРА И АНАЛИЗА ДАННЫХ МЕТЕОРОЛОГИЧЕСКИХ СЕРВИСОВ НА ОСНОВЕ ТЕХНОЛОГИЙ BIG DATA

Аннотация

В докладе рассматривается разработка системы анализа и сравнения точности прогнозов, предоставляемых он-лайн погодными сервисами с помощью технологий Big Data.

Погода и климат давно стали категориями экономическими. Каждый год стихийные бедствия уносят около 250 000 человеческих жизней, размер ущерба, наносимого имуществу, лежит в пределах 50-100 млрд. долларов США. Но мировая статистика показывает: если доверять гидрометеорологической информации и адекватно на нее реагировать, то можно

предотвратить от 30 до 40 % потерь и полностью избежать человеческих жертв [1]. Особенно заметный экономический эффект дает использование метеорологической информации в авиации, энергетике, строительстве, рыболовстве и судоходстве, сельском хозяйстве. Повышение точности прогноза позволяет уменьшить вероятность смертельных случаев во время стихийных бедствий и эффективнее предотвращать экономический и материальный ущерб.

Big Data – совокупность подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объёмов и значительного многообразия для получения воспринимаемых человеком результатов, эффективных в условиях непрерывного прироста, распределения по многочисленным узлам вычислительной сети [2].

Ключевые слова: *Big Data, Apache Hadoop, MapReduce, CDH, Apache Hive, Apache Pig, HDFS, Cloudera.*

Abstract

The report discusses the development of a system of analysis and comparison of the accuracy of forecasts provided by on-line weather services using Big Data technologies.

Weather and climate have long been economic categories. Each year, natural disasters claim about 250,000 lives and damage to property amounts to between \$ 50 billion and \$ 100 billion. However, world statistics show that if you trust hydrometeorological information and respond adequately to it, you can prevent from 30 to 40 % of losses and completely avoid human casualties [1]. The use of meteorological information in aviation, energy, construction, fishing and navigation, agriculture has a particularly noticeable economic effect. Improved forecast accuracy reduces the risk of death in natural disasters and more effectively prevents economic and physical damage.

Big Data – a set of approaches, tools and methods for processing structured and unstructured data of huge volumes and a significant variety to obtain human-perceived results, effective in conditions of continuous growth, distribution across multiple nodes of the computer network [2].

Key words: *Big Data, Apache Hadoop, MapReduce, CDH, Apache Hive, Apache Pig, HDFS, Cloudera.*

Постановка задачи, выбор технологий и средств реализации информационной системы. В ходе выполнения данной работы необходимо собрать данные о прогнозах погоды, предоставляемые различными сервисами.

При выборе сервисов использовались следующие критерии:

- сервис должен иметь API, позволяющий получать данные в виде структурированного файла (XML, JSON, CSV и др.);
- API должен быть свободно распространяемым;
- API должен быть доступен круглосуточно.

Для собранных прогнозных данных требуется провести сравнение с истинными (зафиксированными) показателями погодных условий. Для температурных показателей и силы ветра – рассчитать по каждому сервису среднеквадратичное отклонение прогнозного значения от зафиксированного. Для осадков – рассчитать вероятность ошибки в предсказании осадков. Также можно провести сравнение фактических значений по версии метеорологических сервисов с истинными значениями.

В соответствии с постановкой задачи были выбраны следующие сервисы для получения данных посредством API:

- Метеосервис.py (meteoservice.ru);
- OpenWeatherMap (openweathermap.org);
- World Weather Online (worldweatheronline.com);

– AccuWeather (accuweather.org).

Для реализации проекта выбран фреймворк Apache Hadoop, реализующий методологию Big Data.

Apache Hadoop – это программный проект с открытым исходным кодом, реализующий модель MapReduce и предназначенный для эффективной обработки больших пакетов данных. Вместо одного большого компьютера для обработки и хранения данных Hadoop предлагает использовать для параллельного анализа огромных пакетов данных кластеры на базе стандартного аппаратного обеспечения [4].

Проектирование и реализация программного обеспечения. В составе информационной системы используются следующие компоненты:

- HDFS для хранения получаемых данных;
- Модель MapReduce для обработки данных;
- Apache Pig для очистки и нормализации данных;
- Apache Hive для анализа собираемой информации;
- Cloudera Hue для визуализации данных.

HDFS (англ. Hadoop Distributed File System) – файловая система, предназначенная для хранения файлов больших размеров, поблочно распределённых между узлами вычислительного кластера. Благодаря репликации обеспечивается устойчивость распределённой системы к отказам отдельных узлов. Организация файлов в пространстве имён – традиционная иерархическая: есть корневой каталог, поддерживается вложение каталогов, в одном каталоге могут располагаться и файлы, и другие каталоги [5].

MapReduce – это модель распределённой обработки данных, предложенная компанией Google для обработки больших объёмов данных на компьютерных кластерах [3]. Схема работы MapReduce приведена на рисунке 1.

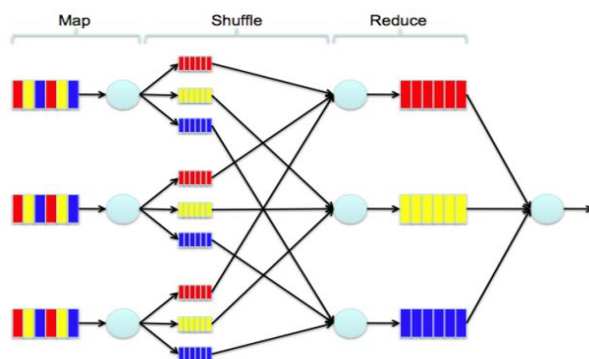


Рис. 1. Схема работы MapReduce

MapReduce предполагает, что данные организованы в виде некоторых записей. Обработка данных происходит в три стадии:

- 1) стадия Map. На этой стадии происходит предобработка данных при помощи функции `map()`, которую определяет пользователь;
- 2) стадия Shuffle. В этой стадии вывод функции `map` разбивается по блокам – каждый блок соответствует одному ключу вывода стадии `map`;

3) стадия Reduce. Каждая «корзина» со значениями, сформированная на стадии shuffle, попадает на вход функции reduce(). Множество всех значений, возвращенных функцией reduce(), является финальным результатом MapReduce-задачи.

Apache Hive – это система хранилища данных для Hadoop. Hive позволяет обобщать, запрашивать и анализировать данные. Запросы Hive создаются на языке запросов HiveQL, который похож на SQL, т.е. Hive представляет из себя инструмент, который превращает SQL-запросы в цепочки MapReduce задач [6].

Apache Pig – это платформа для анализа наборов данных больших объемов, которая состоит из высокоуровневого языка для написания программ анализа данных, а также инфраструктуры, преобразующей код в последовательность MapReduce-задач [7]. В Apache Pig используется язык Pig Latin, который характеризуется следующими ключевыми свойствами:

- простота программирования;
- возможности оптимизации;
- пользователи могут создавать свои собственные функции для специальных целей.

Cloudera Hue – свободно распространяемый программный продукт с открытым исходным кодом, предназначенный для анализа и визуализации данных. Это легковесный веб-сервер, который предоставляет возможность использовать Hadoop напрямую из браузера и который может быть установлен на любую машину Hadoop-кластера.

После выбора компонентов для реализации ИС была разработана структура для хранения и дальнейших манипуляций с полученной информацией. Был создан ряд таблиц для выбранных сервисов:

1. С прогнозами он-лайн сервисов:
 - 1) OpenWeather (openweathermap.org);
 - 2) AccuWeather (accuweather.org);
 - 3) World Weather Online (worldweatheronline.com);
 - 4) Метеосервис.ру (meteoservice.ru).
2. С фактическими (измеренными) значениями от он-лайн сервисов:
 - 1) OpenWeather (openweathermap.org);
 - 2) AccuWeather (accuweather.org);
 - 3) World Weather Online (worldweatheronline.com).

Полученные данные содержат избыточную информацию и не структурированы. В связи с этим возникает необходимость структурирования данных перед записью в БД. Для этого используется высокоуровневый язык обработки Pig Latin.

Скрипт обработки данных принимает на вход исходный файл и схему данных – описание полей фрагмента данных, которые впоследствии будут конвертированы в строку. Далее полученные фрагменты проходят обработку, к ним добавляются дополнительные поля и результаты работы скрипта загружаются в промежуточный файл, который впоследствии будет передан в БД. Блок-схема работы скрипта приведена на рисунке 2.

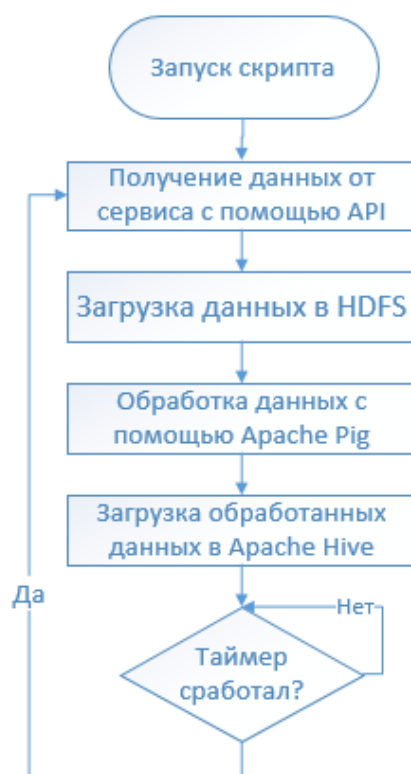


Рис. 2. Блок-схема работы скрипта

После разработки скриптов была определена периодичность их запуска (таблица).

Периодический запуск скриптов производится системной утилитой cron из состава операционной системы CentOS.

На рисунке 3 представлен пример визуализации HiveQL-запроса, выбирающего фактические температуры воздуха 13 апреля 2018 г. по версии сервисов AccuWeather, World Weather Online и OpenWeather.

На основании данного графика можно сделать вывод, что даже фактические данные достаточно сильно (на несколько градусов) отличаются у представленных сервисов.

Таблица

Периодичность запуска скриптов сбора данных

Тип данных	Имя сервиса	Периодичность запуска, часов
Прогноз	OpenWeather	12
	AccuWeather	
	World Weather Online	
	Метеосервис.py	
Фактические данные	OpenWeather	1
	World Weather Online	
	AccuWeather	24

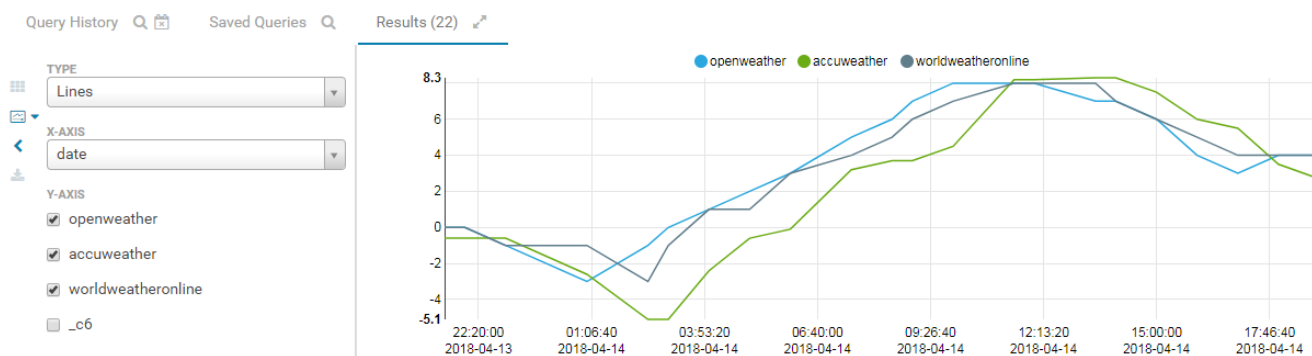


Рис. 3. Визуализация HiveQL-запроса

Возможности разработанной системы. Разработанная информационная система обладает следующими возможностями:

- сбор данных. Анализируемые данные поступают в систему с разной периодичностью;
- отказоустойчивое хранение. Использование HDFS позволяет организовать отказоустойчивое хранение информации без использования дополнительных программных и аппаратных решений;
- очистка и нормализация данных. Собираемые данные представлены различными форматами и нуждаются в предварительной очистке и нормализации.
- анализ и визуализация данных. Для создания выборок данных используется лежащий в основе Apache HIVE язык HiveQL, позволяющий анализировать данные с помощью SQL-запросов. Использование Cloudera Hue позволяет перенести взаимодействие с пользователем из консоли в дружелюбный веб-интерфейс.

Планы по развитию функций системы. При анализе результатов установлено, что «фактические» данные предоставляемые сервисами, различаются. Поэтому для оценки точности прогноза будут использоваться независимый источник данных – архивные данные с метеостанции в г. Екатеринбург.

Также планируется выполнять периодическую проверку точности прогнозов путем сравнения с архивными значениями за выбранный период. Это позволит выбрать сервисы, дающие наиболее точные прогнозы таких параметров, как температура воздуха, сила и направление ветра, вероятность и количество осадков и др.

Выводы. В ходе выполнения данного проекта разработано программное обеспечение, предназначенное для сбора, обработки и представления метеорологических данных, предоставляемых рядом сервисов. Была изучена экосистема Apache Hadoop и произведен выбор соответствующих программных компонентов для реализации поставленной задачи. Созданная информационная система позволяет собирать, обрабатывать и анализировать большие объемы данных используя технологии класса Big Data.

Список использованных источников

1. Стихийные бедствия и техногенные катастрофы. Превентивные меры = Natural Hazards. UnNatural Disasters: The Economics of Effective Prevention. – М.: «Альпина Паблишер», 2012. – 312 с.
2. Большие данные [Электронный ресурс]. Режим доступа: https://ru.wikipedia.org/wiki/Большие_данные – свободный. – Рус.
3. MapReduce [Электронный ресурс]. Режим доступа: <https://ru.wikipedia.org/wiki/MapReduce> – свободный. – Рус.
4. Introduction to Big Data [Электронный ресурс]. Режим доступа: <https://www.coursera.org/learn/big-data-introduction> – свободный. – Англ.
5. Виктор Майер-Шенбергер, Кеннет Кукьер. Большие данные. Революция, которая изменит то, как мы живём, работаем и мыслим = Big Data. A Revolution That Will Transform How We Live, Work, and Think / пер. с англ. Инны Гайдюк. – М.: Манн, Иванов, Фербер, 2014. – 240 с.
6. Apache Pig [Электронный ресурс]. Режим доступа: <https://pig.apache.org/> – свободный. – Англ.
7. Apache Hive [Электронный ресурс]. Режим доступа: <https://hive.apache.org/> – свободный. – Англ.

УДК 004.94

А. А. Першин, Д. А. Чашина, И. А. Гурин

ФГАОУ ВО «Уральский федеральный университет

имени первого Президента России Б.Н. Ельцина», г. Екатеринбург, Россия

РАЗРАБОТКА ИНФОРМАЦИОННО-МОДЕЛИРУЮЩЕЙ СИСТЕМЫ РАЗБАВЛЕНИЯ ПРИМЕСЕЙ В РЕКАХ ПРИ СОСРЕДОТОЧЕННЫХ СТАЦИОНАРНЫХ ВЫПУСКАХ СТОЧНЫХ ВОД

Аннотация

Представлено описание информационно-моделирующей системы, основными функциями которой является расчет разбавления примесей в реках при сосредоточенных стационарных выпусках сточных вод. Отражены основные этапы разработки программного обеспечения: постановка задачи, реализация тестового варианта расчета и проверки методики расчета в электронных таблицах Microsoft Excel; разработка архитектуры информационной системы; проектирование, реализация программного средства (математической библиотеки и пользовательского интерфейса); обработка исключительных ситуаций в программе; разработка системы автоматизированного тестирования для проверки корректности расчетов; создание справочной документации; подготовка дистрибутива; формирование отчета с результатами расчета с возможностью его предварительного просмотра и экспорта во внешние форматы.

Ключевые слова: *сточные воды, примеси в реках, программирование, программное обеспечение, архитектура, интерфейс, математическая библиотека.*